R E P O R T   R E S U M E S

ED 013 537 EM 000 485

THE EFFECTS OF CERTAIN RESPONSE CHARACTERISTICS IN PROGRAMED INSTRUCTION ON ERRORS, RATE OF LEARNING, AND RETENTION.

BY- ORTON, KENNETH D.
NEBRASKA UNIV., LINCOLN

REPORT NUMBER CRP-S-478 PUB DATE 67
REPORT NUMBER BR-5-8433
EDRS PRICE MF-$0.50 HC-$4.04 101P.

DESCRIPTORS- *PROGRAMED INSTRUCTION, *EDUCATIONAL PSYCHOLOGY, *COLLEGE STUDENTS, *TEACHING MACHINES, *RESPONSE MODE, RETENTION, LINEAR PROGRAMING, ACADEMIC PERFORMANCE, EXPERIMENTS, LINCOLN, NEBRASKA

TO INVESTIGATE THE ROLE OF MULTIPLE RESPONSE ALTERNATIVES AS CONSTRAINTS ON THE MEANING OF THE CORRECT RESPONSE, 200 SUBJECTS IN AN INTRODUCTORY EDUCATIONAL PSYCHOLOGY COURSE WERE RANDOMLY ASSIGNED TO 8 TREATMENT GROUPS VARYING ITEM DIFFICULTY, RESPONSE AND SELF-CORRECTION (OVERT V. COVERT), PLUS ONE GROUP FOR USUAL LINEAR PROGRAMING INSTRUCTION. ANALYSIS OF VARIANCE ON IMMEDIATE AND DELAYED ACHIEVEMENT POST-TESTING, USING HALF MULTIPLE CHOICE AND HALF COMPLETION ITEMS SHOWED THE FOLLOWING. EASY ITEMS, CONSTRUCTED RESPONSES AND ERROR CORRECTION RESULTED IN A REDUCED ERROR RATE. DIFFICULT FOILS AND ERROR CORRECTION RESULTED IN SUPERIOR PERFORMANCE ON THE IMMEDIATE COMPLETION TEST, BUT NO FACTORS RESULTED IN DIFFERENTAL PERFORMANCE ON THE DELAYED COMPLETION TEST. THE LINEAR PROGRAM WAS MORE EFFECTIVE FOR ALL CONDITIONS EXCEPT DIFFICULT FOILS AND CORRECTION PROCEDURE. DIFFICULT ITEMS AFFECTED PERFORMANCE ONLY ON THE DELAYED MULTIPLE CHOICE TEST. EASY ITEMS AND THE LINEAR PROGRAM PROMOTED IDENTICAL RESPONSE ERROR REPETITIONS FROM THE IMMEDIATE TO THE DELAYED TEST. RESULTS SUGGEST THAT DIFFICULT, PLAUSIBLE ITEMS COUPLED WITH A CORRECTION PROCEDURE, MAY BE THE MOST EFFECTIVE ADAPTATION OF THE LINEAR PROGRAM. ALL SUBJECTS WHO TOOK THIS PROGRAM OUT-PERFORMED THOSE EXPOSED TO THE LINEAR PROGRAM ON EVERY CRITERION MEASURE. (LH)

A

THE EFFECTS OF CERTAIN RESPONSE CHARACTERISTICS IN PROGRAMED INSTRUCTION ON ERRORS, RATE OF LEARNING, AND RETENTION

COOPERATIVE RESEARCH PROJECT NO S478

KENNETH D. ORTON, Ph.D.

THE UNIVERSITY OF NEBRASKA
LINCOLN, NEBRASKA

1967

The research reported herein was supported by the Cooperative Research Program of the Office of Education U. S. Department of Health, Education, and Welfare.

THE EFFECTS OF CERTAIN RESPONSE CHARACTERISTICS IN PROGRAMED INSTRUCTION ON ERRORS, RATE OF LEARNING, AND RETENTION

Cooperative Research Project No. S478

Kenneth D. Orton, Ph. D.

The University of Nebraska
Lincoln, Nebraska

1967

The research reported herein was supported by the Cooperative Research Program of the Office of Education, U. S. Department of Health, Education, and Welfare.

ACKNOWLEDGMENTS

The support of the Cooperative Research Program of the United States Office of Education in conducting this study is gratefully acknowledged. In addition, the various auxiliary services of the University of Nebraska and more specifically the Department of Educational Psychology and Measurements contributed by providing expertise and encouragement necessary for the completion of the project.

Mr. Roger Bruning, who served as a research assistant, provided the kind of talent which is difficult to secure on projects of this nature. He was invaluable in developing the materials, processing the subjects, analyzing the data, and in general providing critical reactions to the study and eventually to the manuscript. The investigator offers his sincere thanks for these services.

Two students have provided excellent secretarial services during this project. Miss Mary Barton was involved in all phases of the study, and has been particularly valuable in preparing the manuscript. Miss Pamela Lessig has helped in manuscript preparation.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF APPENDICES

CHAPTER I

STATEMENT OF THE PROBLEM

B. F. Skinner's application of his principles of learning to teaching machines and auto-instruction has resulted in considerable debate as well as research on response characteristics associated with these devices. Skinner's views are based on his concept of "shaping" in which gradual changes in response are reinforced. He states that in the delicate process of shaping behavior via program instruction, incorrect responses should be minimized because such responses may be learned (Skinner, 1958). Multiple-choice material which contains plausible wrong responses would tend to promote incorrect responses which would be learned. Skinner is supported in this point of view by Porter (1957).

As a result of this conception both Skinner and Porter have been instrumental in the promotion of programing known as linear programing in which errors in response are held to a minimum by maximizing cueing for each frame within the program. Needless to say, multiple-choice options are virtually eliminated from this type of program.

Pressey (1962) has argued that multiple-choice alternatives should be employed because exposure to wrong alternatives may help to clarify or delimit the meaning of the correct response thereby aiding in the process of developing discriminations. He further suggests that the introduction of plausible alternatives provides a more challenging learning task as compared to a program in which few errors are possible.

However, the programs developed by Pressey are not programs in the same sense as linear programs. The items selected for inclusion are not sequenced nor are they interdependent and interlocking as in linear programs of the type espoused by Skinner. Pressey's approach seems to be a "testing" approach as opposed to a learning approach.

Pressey's different conception of learning has led him to defend the use of multiple-choice items in auto-instruction situations. But some further investigation of the use of multiple-choice responses in the linear programs seems in order.

In addition to the issue stated above, there has been considerable research on the problem of response mode. Should the response be overt and constructed or does covert responding or the indicated response (as in the case of indicating one of the multiple-choice alternatives as correct) function equally well? The evidence to date tends to support some form of overt responding, but differences between overt and indicated responses and overt and constructed responses have not been investigated without confounding the effects of multiple-choice items with the indicated response and the constructed response with a recall type of frame. In this study some subjects will be required to write the alternative they select as correct while others will indicate the alternative they select.

Little information is available on the effects of overt correction of errors in programed instruction. The information that is available suggests that the correction procedure holds some promise. In view of Guthrie's recency principle it would appear that the correction of errors by overtly responding after learning of one's mistakes would tend to counteract the effects of making the error in the first place. If plausible wrong alternatives somehow prompt an error, then overtly responding with the correct answer after learning of that error would seem to be a sound technique for correction.

Finally, the individual and combined effect of the response characteristics on the repetition of errors from immediate tests to delayed tests will be studied. In particular the effect of the plausible wrong alternatives of the multiple-choice frames on errors will be noted. In order for these kinds of analyses to be made, the criterion tests will be composed of selected frames and part-frames of the learning program.

The purpose of this study is to simultaneously investigate the effects of easy versus difficult multiple-choice alternatives, the written response versus the indicated response, and the correction procedure versus the non-correction procedure which allows for additional analysis of interaction effects among these variables.

Objectives

The major objectives of this study are to determine the effects of the response characteristics of difficulty of alternatives, response mode, and correction procedures on a variety of criterion measures. For some oi the above response variables earlier research or intuition suggests the direction of the results when particular criterion measures are employed. In these instances, this investigator can make some statement as to expected outcomes. There are several other effects to be investigated, however, for which statements about expected outcomes are meaningless because of the paucity of data relative to these results. This is particularly true of the double and triple interactions which are by-products of a three-way analysis of variance.

Following are the objectives with some discussion about expected outcome where such discussion seems warranted.

To investigate the effects of the response characteristics of difficulty of alternatives, response mode, and correction procedures and their interactions employing the following criterion measures:

1) Rate of errors in responding during the course of the program.

   By definition one would expect more errors to be associated with the difficult alternatives. One would also expect more errors to be associated with the difficult alternatives, non-correction procedure. The remaining effects are difficult to predict although one would not be too surprised if the correction procedure resulted in fewer errors.

2) Errors on a test of immediate retention employing a recall, constructed response test.

It is predicted that those Ss who are required to construct a response will make fewer errors than those who indicate their response on this criterion measure. Beyond this statement it is difficult to predict the remaining results.

3) Errors on a test of immediate retention employing a multiple-choice test.

It is predicted that the difficult alternatives and the correction procedure will result in the least number of errors with the possibility of the fewest errors being associated with the difficult alternative, correction procedure combination. Past investigations have shown little advantage accruing to a particular mode of response when a multiple-choice type of item is employed.

4) Errors on a test of delayed retention employing the recall, constructed response test.

It appears likely that whatever effects may be evident immediately after the learning situation may, in part, be dissipated over time. A paucity of research data on this question makes prediction difficult.

5) Errors on a test of delayed retention employing a multiple-choice test.

It is not possible to predict the outcome.

6) Items missed on both of the test positions of immediate and delayed retention employing a) the recall, constructed response test, and b) the multiple-choice test.

According to the position taken by Skinner and Holland we would expect the largest number of errors to be associated with difficult alternatives, particularly when employing the multiple-choice test. This statement assumes that more errors will be made on the program with the difficult foils appended and that these errors will endure.

7) Response errors which are repeated on both of the test positions of immediate and delayed retention employing a) the recall, constructed response test, and b) the multiple-choice test.

As per the above comment one would expect a greater number of response error repetitions to be associated with the difficult alternatives.

Related Research

Programed instruction has stimulated a great deal of research of questionable value as well as some of value. In a summary of the research on mode of response Alter and Silverman (1963) suggest that the ease in manipulation of this variable resulted in a large number of investigations, some of which were done rather quickly and poorly with insufficient safeguards including length of program, size of sample, and suitability of the criterion measures. Holland (1965, p. 108) asserts that:

> With a few exceptions the data are collected under poorly controlled conditions. Programed books are often used, and these offer little protection from a variety of distortions of the data including outright cheating by the subjects. Often, the work is done in open classes of 20 to 30 students, in many cases, without the supervision of a trained experimentalist.

The present study was done with the hope that many of the weaknesses of the past research in this area will be eliminated, thus allowing for more precise information.

Multiple-Choice versus the Constructed Response

The most consistent finding of those investigators (Burton and Goldbeck, 1962; Coulson and Silberman, 1960; Fry, 1960; Hough, 1962; Moore and Smith, 1964; Williams, 1965) who have attempted to directly compare multiple-choice response and constructed response on verbal programs has been that of no differences. The major exception was Fry's (1960) study which showed a significant difference in favor of the constructed response tested by recall.

Questions left unanswered by these investigations are many. The majority of these researchers did not specify the procedure for selecting the multiple-choice alternatives. Are we to assume that alternatives were selected to maximize difficulty of the required discrimination? Were synonyms used as logical alternatives? Synonyms may be appropriate alternatives for some programs, but inappropriate for others. For example, it may be that a correct response will have associated with it several other terms, all of which are appropriate in this instance. To force discriminations among these terms would be inappropriate even though the result would undoubtedly be an increase in the number of errors. Conversely, there are instances in which synonymous terms cannot all be considered correct because of the technical nature of the term, and to force discriminations would be appropriate and would also result in a greater number of errors. It is the contention of this investigator that the method of selection of the alternatives will greatly affect the results. Burton and Goldbeck (1962) attempted to research the effects of easy and difficult multiple-choice alternatives, but the study has been rather severely criticized on other grounds, (Holland, 1965) thus rendering the results questionable in value.

The nature of the criterion instrument(s) was not clear in the majority of the above studies. There is little doubt that the selection of items which make up the criterion test is another crucial factor. A congruence between the kinds of discriminations programed and the kinds evaluated would likely result in quite a different outcome than if the criterion test was designed to evaluate the more general aspects of the program. This point is so crucial to the make-up of a sound investigation that to neglect it is to invite criticism.

In addition to these comments on previous studies, one more is in order. The investigators cited above invariably associated the constructed response made with the recall programing frame. Perhaps the effectiveness of the multiple-choice arrangement may be enhanced by requiring subjects to construct the response that they select. Wouldn't a comparison of the recall frame with a multiple-choice frame in which both required a constructed response provide some data of interest? This investigation will attempt to collect some data in answer to this question.

Constructed versus Indicated Response

A large number of research studies have attempted to answer the question of the effect of overt responding as compared to covert responding. Summaries of these kinds of studies have been written by Alter and Silverman (1962) and Holland (1965). Generally, the results have shown no differences, but in a few studies the overt responding technique was superior. Holland, assuming that the overt response is more effective, points out that a post-test advantage for overt responding exists when the answers are contingent on the important content, the error rate is low, and when programs are of sufficient length so as to increase the likelihood that the covert responder will be lulled into not responding at all.

In this investigation the major question in regard to response mode is whether there is a difference between two forms of overt response, that of writing the answer or of indicating the answer. Little evidence is available in regard to this comparison except in those instances in which mode of overt responding is confounded with the issue of multiple-choice versus constructed response.

Correction versus Non-Correction

There is meager research evidence on the effects of correction of a response after feedback. The majority of studies have focused on the effects of feedback without correction of the response. Irion and Briggs (1957) found that correction of the wrong response before proceeding to the next frame resulted in superior performance as compared to a non-correction procedure. However, the correction procedure was inferior to a prompting procedure in which the subject was given a chance to observe the correct response before responding, thus eliminating the possibility of error. In this instance it appears that errors tended to interfere with the acquisition process. Little mention was made of the type of learning program used. It is doubtful that the response was determinate or that the program was anything but a series of rather disconnected items on frames.

Another prompting variation was also superior to the correction procedure. In this approach the subject makes one choice after which he is told that he is right or wrong. If he is wrong the correct response is indicated and the subject then makes the correct response. Here the possibility of making errors is greatly reduced as compared to the correction procedure in which the subject continues to respond until he makes the correct answer. Irion and Briggs speculate that the fact that the stimulus and correct response appeared in close temporal contiguity may be one of the reasons for the success of the prompting techniques.

Suppes and Ginsberg (1962) found that a correction procedure was more effective than a non-correction procedure for children attempting to learn arabic number responses to limited binary number stimuli. The same results did not obtain for adults (Burke, Estes, and Hellyer, 1954).

Glaser (1965) discusses the issue of correction in conjunction with the problem of errors. He states that the concept of correction has received little attention in the literature and is a topic greatly in need of research.

It appears that any investigation of the effects of a correction procedure and learning must, of necessity, be investigated in relationship to a number of additional variables. For example, Kaess and Zeaman (1960) found that the greater the number of plausible alternatives the greater the number of errors. Again, some question is raised about the type of learning program employed which in this case appears to be composed of a series of disconnected items more on the order of a Pressey-type program. Whether these same results would obtain if a linear, low error rate program were employed is a question open to research.

The present study will use a linear program which minimizes error rate and four multiple-choice alternatives in an attempt to investigate the effect of correction and non-correction on programed instruction.

CHAPTER II

PROCEDURE

Subjects

Two hundred subjects from an introductory educational psychology course were randomly assigned into the experimental groups resulting in 25 Ss per cell. Participation in the experimental sessions was a course requirement. The teachers of these classes explained that performance in the experiment would not influence the evaluation of their course performance. The nature of the experiment was briefly explained and Ss were told that the concepts they learned during the experiment would be of value later in the course.

The Program

Six of the first eight sets (1, 2, 3, 4, 7, 8) of Holland and Skinner's Analysis of Behavior (1961) were converted to a multiple-choice format, resulting in a total of 186 frames and 233 responses.[1] The Holland-Skinner program was selected because it is suitable for the available population, and evidence suggests that it is a well-designed program.

A small number of frames in each set were modified so that the central concept under consideration would appear in all frames as the response rather than as part of the stimuli.

---

[1]Permission granted by McGraw-Hill Book Company.

The purpose of these modifications was to facilitate the study of error repetitions by placing the concept itself rather than a functional term in the response position.

Concurrent with the above modifications, three difficult and three easy multiple-choice foils were appended to each frame. These alternatives plus the correct response resulted in four multiple-choice alternatives per frame. The difficult alternatives were derived by selecting synonyms to the correct response and by choosing from among similar psychological concepts which appeared on frames approximately contiguously with the frame in question. It was felt that the choice of "technical" concepts highly related in meaning and encountered close in time to the correct response of a given frame would afford maximal difficulty by forcing discriminations among relatively unfamiliar and yet technical concepts. Examples of the kinds of alternatives developed may be found in Appendix A.

Easy multiple-choice alternatives were generated by selecting non-psychological synonyms to a psychological concept highly dissimilar to the correct response. Thus the task of discriminating among alternatives here would be easily achieved on either or both of the factors of the non-psychological nature of the incorrect alternatives and/or on their dissimilarity of meaning.

The second independent variable under investigation was mode of response. Those subjects completing their programs under the construct mode of response were required to write the response they selected while subjects under the indicate condition were required to indicate their choice by writing the letter corresponding to the response they chose. The programs were identical in format, but different instructions were given to vary the mode of response to be effected by the Ss.

The third and last independent variable under investigation was correction or non-correction of program errors. For the non-correction condition the subject responded to a frame after which the correct response was revealed; he then advanced to the next frame without an overt correction response required in the event of an error. However, for the correction condition, teaching machines were equipped with a second window cut in the display portion of the machine through which S was required to make an overt correction for all response errors made on the program. Corrections were made in the mode of response corresponding to the mode used in making the initial responses. That is, those taking the program under a constructed response condition were required to write out their corrections while those under the indicate condition corrected by writing the letter of the correct choice.

The Experimental Session

From five to 13 subjects participated in an experimental session with all sessions completed within a 17-day period from February 14 to March 5, 1966. Subjects were randomly assigned to treatment conditions in the order of their arrival at the experimental sessions. General instructions were read to all Ss at the beginning of the experimental sessions. They were as follows:

The experiment in which you are participating deals with methods of presenting programed instruction. The material which you will study today deals with some basic principles of learning and is a part of the material which will be covered in your educational psychology course.

Look now to the top of your machines. In the large window across the top, you see the statement, "A doctor taps your knee (patellar tendon) with a rubber hammer to test your ___?___." You complete this statement by indicating your answer in the blank space provided.

Look now at the instructions taped to your machine (PAUSE). These are your specific instructions on how to work through the program. Now, following your instructions, answer the first question in the blank space on the right. (PAUSE).

Now turn either knob and advance the paper about an inch. The right answer will appear from behind the screen while your answer goes up under the glass. Check your answer with the right answer.

Now on some of your machines a second space has been cut on the left. If your answer was wrong AND you have this second space, correct your error according to the instructions on the machine. If you do not have this second space, these last instructions do not apply to you. Are there any questions? (PAUSE).

Once again, follow the instructions taped to your machines exactly in working through the program. If you have any question or your machine isn't working properly, raise your hand. Work at your own speed. You will not all finish at the same time since the programs differ, but you will probably spend about an hour and one-half to two hours on the material. When you finish, there will be a short test over some of the concepts covered in the program. Are there any questions? (PAUSE).

Begin now on item two.

Individual instructions particular to error correction and response mode conditions were taped to the teaching machines appropriately loaded for the combination of conditions desired and Ss' attention directed to these instructions through the general instructions previously presented. Instructions for specific treatment conditions were as follows:

Difficult Construct Correct and Easy Construct Correct

In working through the program, WRITE OUT your answers COMPLETELY in the space on the right. You do not need to write the letter of the answer.

If after checking your answer you see you have made an error, correct it by WRITING OUT COMPLETELY the correct answer in the space on the left.

Difficult Indicate Correct and Easy Indicate Correct

In working through the program, indicate your answer by writing the LETTER of the correct choice in the space on the right. Do not write the words, only indicate your choice by writing A, B, C, or D.

If after checking your answer you see you have made an error, correct it by writing the letter of the correct answer in the space on the left. Again, do not write the words; only indicate the correct answer by writing A, B, C, or D.

Difficult Construct Non-Correct and Easy Construct Non-Correct

In working through the program, WRITE OUT your answers COMPLETELY in the space on the right. You do not need to write the letter of the answer.

Difficult Indicate Non-Correct and Easy Indicate Non-Correct

In working through the program, indicate your answer by writing the LETTER of the correct choice in the space on the right. Do not write the words, only indicate your choice by writing A, B, C, or D.

Arrows led from the instructions to the appropriate space on the machine. Additionally, for the four conditions involving non-correction, machines were not adapted to permit correction and no reference to the need for correction was made in these individual instructions.

A starting time was recorded for each S at this point in the session. Twenty minutes after the beginning of the session exhibits associated with Sets 3, 4, and 8 were handed out to each subject and the following instructions were read to the group:

Later in the program, you will be instructed by the program to refer to these materials. Do not read them now, only when you are told to read them in the program.

As each S finished the program, the time taken to complete the program was recorded, and the S was required to respond to 54 test items composed of 27 completion items and 27 multiple-choice items. The order of administration of the two types of items was systematically varied for each S with the order of administration remaining the same for a given subject for both immediate and delayed testing.

Criterion Measures

Time taken to complete the program, number of errors on the program, and errors made on the two-part test mentioned previously were the criterion measures employed. The latter measure was administered immediately after the exposure to the program and again approximately two weeks later. The delayed test varied as much as two days per subject because all subjects were tested the second time within their regular classroom session. Since this variation was consistent for all experimental conditions, it was felt that the results of this study were not biased.

The test items for the two types of criterion tests were composed of parts of program frames considered most appropriate as a test of each major concept presented. Part frames were used in order to eliminate cues designed to evoke the correct response. The completion items were taken directly from the linear program without any further modification. The multiple-choice items were constructed by using the same part frames as previously discussed, but with difficult multiple-choice alternatives appended which were different from those used on the program. For the 27 program frames employed later as test items, a pool of six difficult multiple-choice alternatives were selected. Alternatives were then paired according to degree of difficulty and members of each pair randomly assigned to the program with the remaining member then assigned to the test.

An attempt to analyze the effects of errors made during the program on errors made later on the criterion tests failed for lack of data. Because of the low error rate of the program and the fact that approximately 25 percent of the responses required during the program were included on the criterion tests there were many instances when the Ss did not repeat any of the errors on the tests that were made on the program.

However, analyses of error repetition from the immediate to the delayed tests were completed. That is, cases in which the S missed the items on both the immediate and delayed tests, but made a different response each time, were recorded and used as a criterion measure. And finally, those cases in which the response was in error on the immediate tests and the same, wrong, response was repeated on the delayed tests were also recorded and used as a criterion measure.

Treatment of the Data

Combination of the three major independent variables resulted in the following treatment conditions:

1. Difficult Alternatives-Constructed Responses-Correction (DCC)

2. Difficult Alternatives-Constructed Responses-Noncorrection (DCNC)

3. Difficult Alternatives-Indicated Responses-Correction (DIC)

4. Difficult Alternatives-Indicated Responses-Noncorrection (DINC)

5. Easy Alternatives-Constructed Responses-Correction (ECC)

6. Easy Alternatives-Constructed Responses-Noncorrection (ECNC)

7. Easy Alternatives-Indicated Responses-Correction (EIC)

8. Easy Alternatives-Indicated Responses-Noncorrection (EINC)

All criterion measures previously mentioned were serially employed with a 2 x 2 x 2 factorial analysis of variance statistical design. Main and interaction effects resulted from these analyses and simple effects were tested where appropriate.

CHAPTER III

RESULTS

The same sets of the Holland-Skinner program that were used for the experimental groups were administered in their usual format, that of the recall-construct type with feedback but no correction, to seventeen students. The purpose of this adjunct to the proposed research was to provide a basis for comparing a usual linear program with the various modified formats used in this study, and in particular to compare the usual linear program with the condition in which the students constructed answers and did not correct (DCNC). The latter condition was similar to the usual linear program in every way except for the difficult multiple-choice foils appended to each response position within the frame. A comparison between these two programs should generate data which is germane to the issue of the effect of multiple-choice responses on error production on the program and on the criterion tests as well.

Descriptive Data

The mean proportion of errors on all experimental conditions and on the usual linear program is presented in Table 1. As expected, those Ss exposed to the easy alternatives made the fewest errors on the program. The Ss exposed to the

TABLE 1

PROPORTION OF ERRORS TO TOTAL RESPONSES
ON THE PROGRAM AND ON IMMEDIATE AND DELAYED COMPLETION
AND MULTIPLE-CHOICE CRITERION TESTS

TREATMENT CONDITION	PROGRAM	IMMEDIATE COMPLETION	DELAYED COMPLETION	IMMEDIATE MULT.CHOICE	DELAYED MULT.CHOICE
DCC	.086	.186	.357	.046	.081
DIC	.129	.261	.420	.084	.141
DCNC	.133	.310	.449	.104	.144
DINC	.162	.343	.461	.099	.124
ECC	.015	.319	.469	.109	.143
EIC	.020	.341	.410	.080	.156
ECNC	.014	.350	.450	.101	.161
EINC	.020	.374	.464	.117	.183
RECALL-CONSTRUCT	.192	.273	.370	.102	.111

difficult alternatives were next in error production with the constructed response and correction being associated with the fewest errors within that category. Finally, Ss responding to the usual linear program made the greatest number of errors in comparison to all conditions. The addition of multiple-choice alternatives does not appear to have an adverse effect on error production as compared to the usual linear program.

Scrutiny of the test results reveals quite different findings. It is noted that the Ss under the DCC and DIC conditions made the fewest errors on the completion test administered immediately with the usual linear program a close third. All other conditions yielded a larger number of errors on the completion test administered immediately. It is also clear that the indicate and non-correction conditions are consistently associated with the larger number of errors regardless of the difficulty level of the alternatives.

The results of the delayed-completion test show that only the DCC condition is superior to the usual linear approach, and that all other conditions generated a relatively greater number of errors. The results are somewhat more variable on this criterion measure.

Again, the DCC condition yielded the smallest number of errors using the immediate-multiple-choice test as the criterion with two conditions in which the response was indicated and corrected (DIC and EIC) next. These results are somewhat less variable than those noted previously.

Finally, using the delayed multiple-choice test as the criterion it is noted that the DCC condition is again the most effective in terms of proportion of errors with the usual linear program second.

Analysis

As was mentioned earlier, the data was analyzed by a 2 x 2 x 2 factorial analysis of variance design. This analysis was repeated nine times, once for each criterion measure. A follow-up analysis of significant main effects is reported for each criterion measure.

Program Errors

Table 2 shows the analysis of variance for program errors as well as the means and standard deviations for each of the treatment conditions. As would be expected, the variable of difficulty is highly significant with variables of response mode and correction being significant at the .05 level of confidence. In addition, one two-way interaction, that of difficulty x correction, was significant at the .05 level. Perhaps the most surprising finding is that the condition of the response mode was significant since both forms of responding can be classified as overt forms.

Inspection of the means and standard deviations suggests that the assumptions of normality and homogeneity of variance can be questioned. In this study the finding of differential error rates associated with the several conditions only confirms

TABLE 2

ANALYSIS OF VARIANCE AND
MEANS AND STANDARD DEVIATIONS USING
TOTAL PROGRAM ERRORS
AS THE CRITERION MEASURE

SOURCE OF VARIATION	SUMS OF SQUARES	df	MEAN SQUARES	F
DIFFICULTY	33282.00	1	33282.00	170.614 **
RESPONSE MODE	1190.72	1	1190.72	6.104 *
DIFFICULTY X MODE	655.22	1	655.22	3.359
CORRECTION	1039.68	1	1039.68	5.330 *
DIFFICULTY X CORRECT	1142.42	1	1142.42	5.856 *
MODE X CORRECT	27.38	1	27.38	0.140
DIFFICULTY X MODE X CORRECT	38.72	1	38.72	0.198
MODE WITHIN	37553.84	192	195.59	
TOTAL	74829.98	199		

* $p < .05$
** $p < .01$

MEANS & STANDARD DEVIATIONS OF CELLS

	MEAN	S.D.		MEAN	S.D.
DCC	20.16	9.37	ECC	3.64	3.26
DIC	30.28	19.10	EIC	4.76	4.88
DCNC	31.12	18.75	ECNC	3.28	3.45
DINC	38.00	26.25	EINC	4.68	4.61

that the conditions, arranged on *a priori* grounds, do in fact affect error rate. In particular this is apparent for the dimension of difficulty. A relatively exact test of the effect of these conditions on error rate is not seen to be necessary. In addition, evidence in Lindquist (1953) and Scheffe (1959) suggests that the tests under these conditions will be fairly accurate providing there are equal cell *N*'s, which is the case in this study.

An attempt was made to determine which combination of variables was associated with the best performance or lowest error rate on the program by comparing specified combinations of cells. For every major variable that was found to be significant beyond the .05 level of significance, eight cell comparisons were made. For example, the following comparisons were made for the dimension of *difficulty* since that variable was significant:

A. Comparisons in which the data was collapsed over the correction dimension.

1. Easy-Construct vs. Difficult-Construct
2. Easy-Indicate vs. Difficult-Indicate

B. Comparisons in which the data was collapsed over the response mode dimension.

3. Easy-Correct vs. Difficult-Correct
4. Easy-Noncorrect vs. Difficult-Noncorrect

C. Comparisons of the original treatment combinations.

5. Easy-Construct-Correct vs. Difficult-Construct-Correct
6. Easy-Indicate-Correct vs. Difficult-Indicate-Correct
7. Easy-Construct-Noncorrect vs. Difficult-Construct-Noncorrect
8. Easy-Indicate-Noncorrect vs. Difficult-Indicate-Noncorrect

Every one of these eight comparisons shows a highly significant (.01) difference in favor of the difficult alternatives of the difficulty dimension. Again, this is expected since the alternatives were deliberately selected to affect error rate.

Later reports of the follow-ups of the significant major variables will include only those comparisons which yield significant results at the .05 level or better.

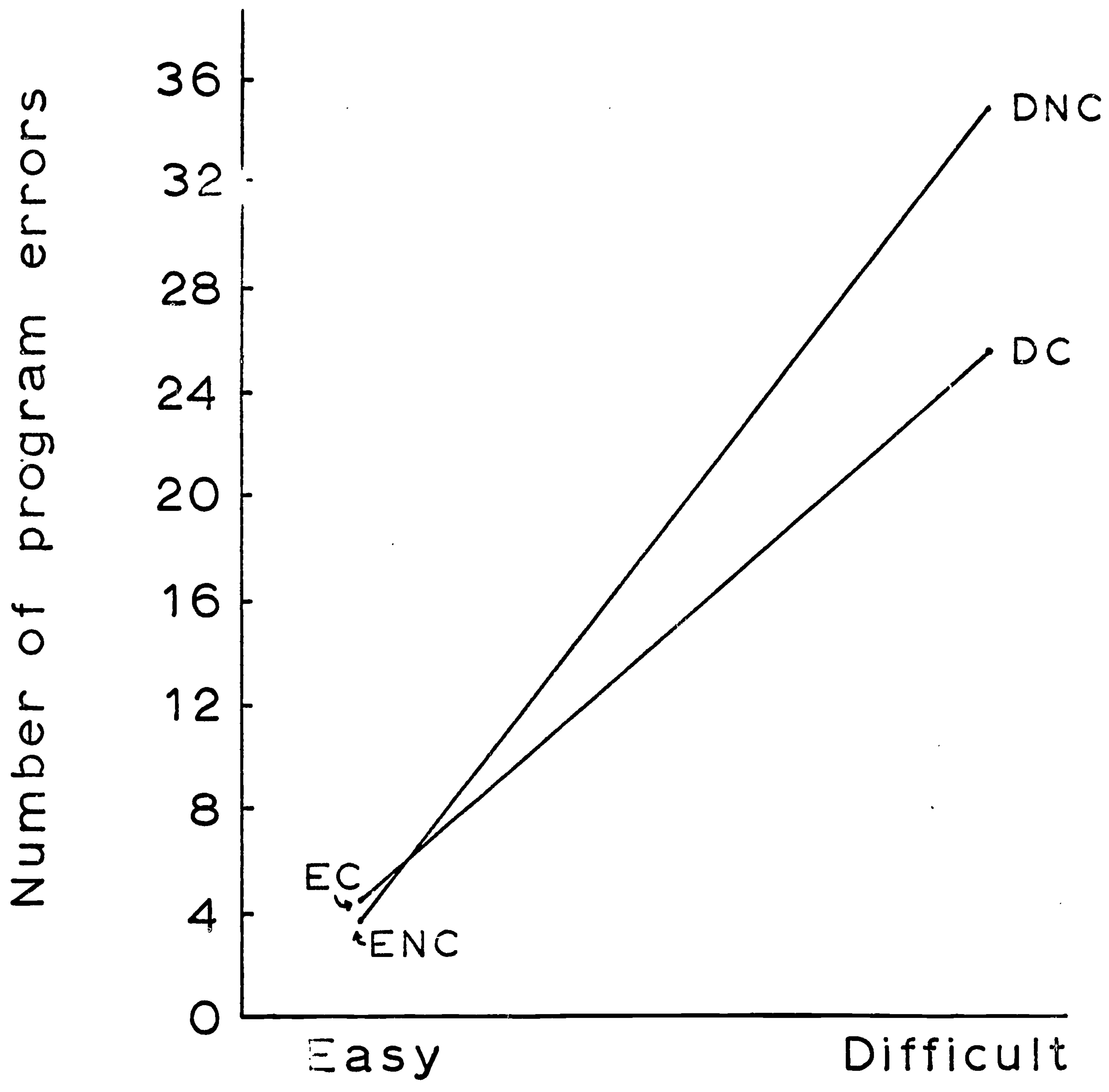
Only two comparisons of the possible eight are significant in the follow-up of the significant major variable of response mode. The comparison of Difficult-Construct vs. Difficult-Indicate when the data were collapsed over the correction dimension and DCC vs. DIC were both significant. It can be seen that the constructed response will result in fewer errors than the indicated response, especially when the alternatives are difficult.

The correction dimension was a significant factor in reducing programing errors. The follow-up analyses showed that the Difficult-Correct vs. Difficult-Noncorrect conditions were significantly different when the data were collapsed over the response mode dimension. The DCC vs. DCNC was also significant. These results suggest that correction is an important factor in the reduction of errors when the alternatives are difficult and the constructed response is employed.

Finally, the significant interaction effect between the difficult and correction dimensions is presented graphically in Figure 1. It can be seen that the factor of correction does not much affect the error rate when easy alternatives are employed. The reverse is true for the difficult alternatives.

FIGURE 1

PLOT OF COLLAPSED MEANS TO SHOW INTERACTION



In summary of the results using program errors as the criterion measure, easy alternatives, constructed response, and correction of errors resulted in the fewest errors, but with the constructed response and correction of errors functioning best in conjunction with difficult alternatives.

Immediate Completion Test

The analysis of variance for the immediate completion test and the means and standard deviations for each of the treatment conditions is presented in Table 3. Significant results were obtained for the dimension of difficulty at the .01 level and for the dimension of correction at the .05 level. The means indicate that the difficult alternatives and correction activity resulted in the least number of errors on the completion test administered immediately.

Follow-up comparisons of the type described before were executed with each of the significant dimensions. For the _difficulty_ dimension, significant differences (at the .05 level) were found between Difficult-Construct and Easy-Construct when the data were collapsed over the correction dimension, and between Difficult-Correct and Easy-Correct when the data were collapsed over the response mode dimension. In addition, the DCC vs. ECC comparison was also significant. It is clear that the more difficult alternatives resulted in fewer errors on the completion test administered immediately, especially when the difficult alternatives were combined with the constructed response and/or the correction procedure.

TABLE 3

ANALYSIS OF VARIANCE AND
MEANS AND STANDARD DEVIATIONS USING
COMPLETION TEST ERRORS
AS THE CRITERION MEASURE

SOURCE OF VARIATION	SUMS OF SQUARES	df	MEAN SQUARES	F
DIFFICULTY	198.00500	1	198.00500	7.418 **
RESPONSE MODE	59.40500	1	59.40500	2.225
DIFFICULTY X MODE	9.24500	1	9.24500	0.346
CORRECTION	178.60500	1	178.60500	6.691 *
DIFFICULTY X CORRECT	49.00500	1	49.00500	1.836
MODE X CORRECT	4.20500	1	4.20500	0.158
DIFFICULTY X MODE X CORRECT	4.80500	1	4.80500	0.180
MODE WITHIN	5125.11984	192	26.69333	
TOTAL	5628.39484	199		

* $p < .05$
** $p < .01$

MEANS & STANDARD DEVIATIONS OF CELLS

	MEAN	S.D.		MEAN	S.D.
DCC	5.20000	3.20156	ECC	8.92000	5.02427
DIC	7.32000	4.96421	EIC	9.56000	5.35475
DCNC	8.68000	5.80747	ECNC	9.80000	4.51848
DINC	9.60000	6.23832	EINC	10.48000	5.62820

Similarly, follow-up analyses were made with the correction dimension. It was found that Difficult-Correction conditions made significantly fewer errors than the Difficult-Noncorrection condition when both of these conditions were collapsed over the dimension of response mode, and the Construct-Correct condition resulted in significantly fewer errors than the Construct-Noncorrect condition when the difficulty dimension was ignored. DCC vs. DCNC was significant among all comparisons of interest. These results suggest that the correction procedure was of value in error reduction when combined with difficult alternatives and when the response mode is the constructed response.

Summarizing the results using the immediate completion test as a criterion measure, it was found that difficult alternatives and a correction procedure resulted in the fewest test errors. These factors functioned best in combination with one another and in conjunction with the constructed response.

Delayed Completion Test

The analysis of variance for the delayed completion test and the means and standard deviations for each of the treatment conditions are presented in Table 4. None of the effects are significant although it is noted that the DCC condition again resulted in the best performance. It appears that the effects of the variables of difficulty and correction, which were significant on the immediate completion test criterion, are somewhat diminished over time.

TABLE 4

ANALYSIS OF VARIANCE AND
MEANS AND STANDARD DEVIATIONS USING
DELAYED COMPLETION TEST ERRORS
AS THE CRITERION MEASURE

SOURCE OF VARIATION	SUMS OF SQUARES	df	MEAN SQUARES	F
DIFFICULTY	27.38000	1	27.38000	1.183
RESPONSE MODE	2.42000	1	2.42000	----
DIFFICULTY X MODE	35.28000	1	35.28000	1.524
CORRECTION	69.62000	1	69.62000	3.007
DIFFICULTY X CORRECT	23.12000	1	23.12000	----
MODE X CORRECT	1.28000	1	1.28000	----
DIFFICULTY X MODE X CORRECT	36.98000	1	36.98000	1.597
MODE WITHIN	4445.43984	192	23.15333	
TOTAL	4641.51984	199		

MEANS & STANDARD DEVIATIONS OF CELLS

	MEAN	S.D.		MEAN	S.D.
DCC	10.00000	4.11229	ECC	13.12000	5.26245
DIC	11.76000	4.86724	EIC	11.48000	5.10816
DCNC	12.56000	4.77040	ECNC	12.60000	3.84057
DINC	12.92000	4.78644	EINC	13.00000	5.51513

Immediate Multiple-Choice Test

The analysis of variance and descriptive data for the immediate multiple-choice test criterion are presented in Table 5. None of the effects are significant at the .05 level, but the dimension of correction was very close (3.88 where 3.90 needed to be significant at the .05 level). Once again it is noted that the DCC condition resulted in the least number of errors which is not expected when a multiple-choice test is the criterion. The investigator felt that the indicated response would do at least well as the constructed response when this criterion test was employed.

Delayed Multiple-Choice Test

The analyses for this delayed criterion measure and descriptive data are presented in Table 6. The only factor which was significant was that of difficulty at the .05 level. Follow-up data on the variable of difficulty within the several comparisons previously specified failed to yield any significant results although all comparisons show the least number of errors to be associated with the difficult alternatives. Once again, response mode is not the powerful factor one would expect it to be when this type of criterion is employed.

Percent of Errors Made on Immediate Completion Test Which Were Repeated on Delayed Completion Test

The data presented up to this point is based on the number of errors made on either the program or immediate or delayed response to criterion tests, and has shown the dimensions

TABLE 5

ANALYSIS OF VARIANCE AND
MEANS AND STANDARD DEVIATIONS USING
IMMEDIATE MULTIPLE-CHOICE TEST ERRORS
AS THE CRITERION MEASURE

SOURCE OF VARIATION	SUMS OF SQUARES	df	MEAN SQUARES	F
DIFFICULTY	13.52000	1	13.52000	2.026
RESPONSE MODE	0.98000	1	0.98000	0.147
DIFFICULTY X MODE	5.12000	1	5.12000	0.767
CORRECTION	25.92000	1	25.92000	3.88
DIFFICULTY X CORRECT	4.50000	1	4.50000	0.674
MODE X CORRECT	0.00000	1	0.00000	-----
DIFFICULTY X MODE X CORRECT	19.22000	1	19.22000	2.880
MODE WITHIN		192	6.67250	
TOTAL	1350.37997	199		

MEANS & STANDARD DEVIATIONS OF CELLS

	MEAN	S.D.		MEAN	S.D.
DCC	1.28000	1.92614	ECC	3.04000	2.97881
DIC	2.36000	2.95635	EIC	2.24000	2.24128
DCNC	2.92000	3.08113	ECNC	2.84000	2.33952
DINC	2.76000	2.16564	EINC	3.28000	2.71600

TABLE 6

ANALYSIS OF VARIANCE AND
MEANS AND STANDARD DEVIATIONS USING
DELAYED MULTIPLE-CHOICE TEST ERRORS
AS THE CRITERION MEASURE

SOURCE OF VARIATION	SUMS OF SQUARES	df	MEAN SQUARES	F
DIFFICULTY	56.18000	1	56.18000	5.1076 *
RESPONSE MODE	13.52000	1	13.52000	1.229
DIFFICULTY X MODE	0.08000	1	0.08000	----
CORRECTION	20.48000	1	20.48000	1.862
DIFFICULTY X CORRECT	0.00000	1	0.00000	----
MODE X CORRECT	12.50000	1	12.50000	1.136
DIFFICULTY X MODE X CORRECT	19.22000	1	19.22000	1.747
MODE WITHIN	2111.83998	192	10.9917	
TOTAL	2233.81998	199		

* $p < .05$

MEANS & STANDARD DEVIATIONS OF CELLS

	MEAN	S.D.		MEAN	S.D.
DCC	2.28000	2.80654	ECC	4.00000	3.62859
DIC	3.96000	3.28481	EIC	4.36000	3.23883
DCNC	4.04000	2.86473	ECNC	4.52000	3.60694
DINC	3.48000	2.34734	EINC	5.12000	4.35239

of difficulty and correction to be most often associated with the best performances. The question raised here is, "Will the dimension of difficulty be associated with greater percentage of repeated errors from the immediate test to the delayed test?" That is, what percent of the time will a person who misses the item on the immediate test also miss the item on the delayed test? Since the tests employed in this study are actually selected frames and part frames of the program, with equally difficult foils appended, one would expect that errors induced by difficult choices on the program would tend to persist and be repeated on later tests. For easy choices, it is noted that fewer errors were made on the program, but a greater number of errors were made on the first tests as compared to those who were exposed to difficult choices. One can only guess whether or not the same items will be missed on an identical test administered at a later date.

The analysis and descriptive data for item error repetition for both the completion test and the multiple-choice test are presented in Tables 7 and 8. None of the results presented is significant, but some tendencies were found when the completion test is employed as the criterion measure. It is noted that a smaller percent of error repetition is associated with the difficult alternatives combined with constructed response and the correction condition (DCC). This finding, though not significant, is counter to the idea that difficult alternatives will result in persistence of a particular type of error.

TABLE 7

ANALYSIS OF VARIANCE AND
MEANS AND STANDARD DEVIATIONS USING
ITEM ERROR REPETITION ON COMPLETION TESTS
AS THE CRITERION MEASURE

SOURCE OF VARIATION	SUMS OF SQUARES	df	MEAN SQUARES	F
DIFFICULTY	0.22526	1	0.22526	3.391
RESPONSE MODE	0.00000	1	0.00000	----
DIFFICULTY X MODE	0.00445	1	0.00445	----
CORRECTION	0.02632	1	0.02632	----
DIFFICULTY X CORRECT	0.04304	1	0.04304	----
MODE X CORRECT	0.15466	1	0.15466	2.328
DIFFICULTY X MODE X CORRECT	0.20241	1	0.20241	3.047
MODE WITHIN	12.75460	192	0.06643	
TOTAL	13.41074	199		

MEANS & STANDARD DEVIATIONS OF CELLS

	MEAN	S.D.		MEAN	S.D.
DCC	0.59186	0.35335	ECC	0.76138	0.26837
DIC	0.72049	0.22733	EIC	0.74389	0.21042
DCNC	0.76338	0.23431	ECNC	0.74697	0.21889
DINC	0.65353	0.31206	EINC	0.74551	0.19599

TABLE 8

ANALYSIS OF VARIANCE AND
MEANS AND STANDARD DEVIATIONS USING
ITEM ERROR REPETITION ON MULTIPLE-CHOICE TESTS
AS THE CRITERION MEASURE

SOURCE OF VARIATION	SUMS OF SQUARES	df	MEAN SQUARES	F
DIFFICULTY	0.06836	1	0.06836	----
RESPONSE MODE	0.01288	1	0.01288	----
DIFFICULTY X MODE	0.70277	1	0.70277	1.825
CORRECTION	0.14405	1	0.14405	----
DIFFICULTY X CORRECT	0.18311	1	0.18311	----
MODE X CORRECT	0.00110	1	0.0110	----
DIFFICULTY X MODE X CORRECT	0.00527	1	0.00527	----
MODE WITHIN	73.91668	192	0.38498	---
TOTAL	75.03422	199		

MEANS & STANDARD DEVIATIONS OF CELLS

	MEAN	S.D.		MEAN	S.D.
DCC	0.23111	1.00560	ECC	0.45743	0.39501
DIC	0.38067	0.38645	EIC	0.34933	0.40278
DCNC	0.36025	0.39552	ECNC	0.44500	0.36418
DINC	0.47990	1.09367	EINC	0.34808	0.34034

Percent of Response Errors Identically Repeated on Both Completion and Multiple-Choice Tests

If S made a particular response error when either test was immediately administered, did he also make that same response error on that same item when the test was readministered at a later date? The data reported in Tables 9 and 10 are based on the percent of identical response error repetition from the immediate completion test to the delayed completion test and from the immediate multiple-choice test to the delayed multiple-choice test. Only one significant effect is reported, that of the difficulty dimension when the completion test is employed as the criterion measure. Inspection of the means shows a smaller percent of identical response error repetitions consistently associated with the difficult alternatives. That is, a greater proportion of identical response error repetitions are associated with easy alternatives on the program rather than difficult alternatives. This finding indirectly refutes the contention of Skinner (1958) and Porter (1957) that one tends to learn the errors made on the program and that these errors will persist.

Follow-up analyses show that a greater incidence of identical response error repetition is associated with the easy alternatives for all comparisons between difficult alternatives and easy alternatives, but none of these comparisons attained significance. Apparently, these consistent differences resulted in a significant main effect for the difficulty dimension.

TABLE 9

ANALYSIS OF VARIANCE AND
MEANS AND STANDARD DEVIATIONS USING
IDENTICAL RESPONSE ERROR REPETITION ON COMPLETION TESTS
AS THE CRITERION MEASURE

SOURCE OF VARIATION	SUMS OF SQUARES	df	MEAN SQUARES	F
DIFFICULTY	0.22367	1	0.22367	4.024 *
RESPONSE MODE	0.01783	1	0.01783	----
DIFFICULTY X MODE	0.00052	1	0.00052	----
CORRECTION	0.01262	1	0.01262	----
DIFFICULTY X CORRECT	0.00671	1	0.00671	----
MODE X CORRECT	0.15653	1	0.15653	2.816
DIFFICULTY X MODE X CORRECT	0.00753	1	0.00753	----
MODE WITHIN	10.67093	192	0.05558	----
TOTAL	11.09634	199		

* $p < .05$

MEANS & STANDARD DEVIATIONS OF CELLS

	MEAN	S.D.		MEAN	S.D.
DCC	0.30400	0.26107	ECC	0.39153	0.27171
DIC	0.35012	0.25242	EIC	0.41954	0.23197
DCNC	0.36792	0.25042	ECNC	0.40773	0.20101
DINC	0.27760	0.19772	EINC	0.34838	0.20714

TABLE 10

ANALYSIS OF VARIANCE AND
MEANS AND STANDARD DEVIATIONS USING
IDENTICAL RESPONSE ERROR REPETITION ON MULTIPLE-CHOICE TESTS
AS THE CRITERION MEASURE

SOURCE OF VARIATION	SUMS OF SQUARES	df	MEAN SQUARES	F
DIFFICULTY	0.22599	1	0.22599	1.62
RESPONSE MODE	0.01416	1	0.01416	----
DIFFICULTY X MODE	0.40030	1	0.40030	2.874
CORRECTION	0.20597	1	0.20597	1.479
DIFFICULTY X CORRECT	0.32940	1	0.32940	2.365
MODE X CORRECT	0.16245	1	0.16245	1.166
DIFFICULTY X MODE X CORRECT	0.03274	1	0.03274	----
MODE WITHIN	26.74647	192	0.13930	
TOTAL	28.11748	199		

MEANS & STANDARD DEVIATIONS OF CELLS

	MEAN	S.D.		MEAN	S.D.
DCC	0.05778	0.21545	ECC	0.32124	0.36114
DIC	0.24667	0.29214	EIC	0.28000	0.35013
DCNC	0.28571	0.37538	ECNC	0.33567	0.34797
DINC	0.30943	0.62453	EINC	0.23160	0.27861

Inspection of the means and standard deviations on Table 9 shows a general decline in the percent of identical response error repetition with the greatest reduction resulting from the DCC condition. However, it is also noted that seven of eight standard deviations are considerably larger, which probably accounts for the findings of nc significance for any of the main effects.

Comparison of the Usual Linear Program with the Difficult-Construct-Noncorrect Condition

It was previously stated that the usual linear program (RCNC) was administered and the same criterion measures taken in order to provide a basis for comparison with the other conditions of the study and to specifically compare these results with the results of the difficult-construct-noncorrect (DCNC) condition. The results are presented in Table 11. The only significant result obtained by these comparisons was on the criterion measure of program errors. The DCNC condition resulted in significantly fewer program errors. Two other comparisons, that of time taken to complete the program and the percent of identical response error repetitions, favored the DCNC condition. All others favored the RCNC condition. The Ss under the RCNC condition consistently performed better than the Ss under the DCNC condition, but not significantly so.

TABLE 11

MEANS, VARIANCES, AND T-TESTS
ON COMPARISONS OF THE RCNC PROGRAM
WITH THE DCNC PROGRAM

VARIABLE	PROGRAM	MEAN	VARIANCE	T-TEST
TOTAL PROGRAM ERRORS	RCNC DCNC	44.88 31.12	406.10 351.52	-2.26*
NO. ERRORS IMMEDIATE COMPLETION TEST	RCNC DCNC	7.65 8.68	27.05 33.72	-0.588
NO. ERRORS IMMEDIATE MULTIPLE-CHOICE TEST	RCNC DCNC	2.88 2.92	5.87 9.49	-0.045
NO. ERRORS DELAYED COMPLETION TEST	RCNC DCNC	10.35 12.56	26.70 22.76	-1.42
NO. ERRORS DELAYED MULTIPLE-CHOICE TEST	RCNC DCNC	3.12 4.04	4.46 8.21	-1.13
ITEM ERROR REPETITIONS ON COMPLETION TESTS	RCNC DCNC	.705 .763	.063 .055	-0.733
IDENTICAL RESPONSE ERROR REPETITIONS ON COMPLETION TESTS	RCNC DCNC	.384 .367	.091 .063	-0.200
ITEM ERROR REPETITIONS ON MULTIPLE-CHOICE TESTS	RCNC DCNC	.241 .360	.113 .156	-1.01
IDENTICAL RESPONSE ERROR REPETITIONS ON MULTIPLE-CHOICE TESTS	RCNC DCNC	.183 .286	.068 .141	-0.990
TIME TAKEN TO COMPLETE PROGRAM (IN MINUTES)	RCNC DCNC	105.12 100.52	214.34 378.76	-0.829

$*p < .05$

Additional Analyses

In the course of conducting research the investigator usually has some new "insights" or sees the project in a somewhat different light than formerly. Such was the case in this investigation. On the following pages a brief discussion of the procedures and findings of selected analyses will be presented.

A Comparison of Performance on Two Forms of Multiple-Choice Tests Under the Difficult-Constructed Response-Noncorrection Condition

The multiple-choice test used in this study was developed by selecting "critical" frames and part-frames from the program and appending difficult multiple-choice foils to those frames. However, these difficult foils were not the same as those on the modified program and therefore may have interfered with the response of those Ss who were previously exposed to yet other difficult alternatives. Or would having the same identical foils on both the program and the test result in more errors? In order to clarify the issue a group of Ss numbering 25 were administered the difficult alternative-constructed response-noncorrection (DCNC) program. It is recalled that one other group responded to this same program. The major difference between this group and the group who earlier took this program was that this group was administered both the old completion and multiple-choice tests with the difficult alternatives of the program appended to each frame. Now if an S makes an error on the program, will that error tend to persist through both the immediate and delayed tests because the cues are identical from the learning to the test situation? Will the error rate and error repetition be greater for the DCNC condition which takes the test with the identical foils or will Ss who take the test with different difficult foils be more prone to error production and error repetition? Data bearing on these questions is presented in Table 12. Within this table the original group is labeled DCNC, and the group with the tests which have foils identical to those on the learning program are labeled DCNC-Identical Foils Test (IFT). The comparisons of interest are those associated with the multiple-choice tests as opposed to the completion tests. In every comparison the group which took the multiple-choice test with the same foils as appeared on the program either made more errors than the original DCNC group or made significantly more error repetitions. Apparently, re-presenting the identical cue situation as was encountered on the program results in greater error production and error repetition as compared to a test situation in which the alternatives to the correct response were difficult but different.

Analysis of Program Errors

A number of errors made on the learning program could be repeated on the tests that followed since the tests consisted of a sampling of the frames of the program. However, the number of possible error repetitions per S was relatively small. The number of identical response error repetitions, item error repetitions, and item corrections from the program to the test are reported in Tables 13 and 14 for the program conditions of RCNC,

TABLE 12

MEANS, VARIANCES, AND T-TESTS
ON COMPARISONS OF THE DCNC-IFT GROUP
WITH THE ORIGINAL DCNC GROUP

VARIABLE	PROGRAM-TEST	MEANS	VARIANCES	T-TEST
TOTAL PROGRAM ERRORS	DCNC-IFT	34.08	407.11	-0.531
	DCNC	31.12	351.52	-0.531
NO. ERRORS IMMEDIATE COMPLETION TEST (DCNC-IFT had multiple-choice here)	DCNC-IFT (M-C)	5.04	12.99	
	DCNC (Comp.)	8.68	33.72	-2.66*
NO. ERRORS IMMEDIATE MULTIPLE-CHOICE TEST	DCNC-IFT	4.84	9.89	
	DCNC	2.92	9.49	-2.18*
NO. ERRORS DELAYED COMPLETION TEST (DCNC-IFT had multiple-choice here)	DCNC-IFT (M-C)	5.60	12.24	
	DCNC (Comp.)	12.56	22.76	-5.88**
NO. ERRORS DELAYED MULTIPLE-CHOICE TEST	DCNC-IFT	6.72	10.68	
	DCNC	4.04	8.21	-3.08**
PERCENTAGE OF ITEM ERROR REPETITIONS ON MULTIPLE-CHOICE TESTS	DCNC-IFT	.652	.065	
	DCNC	.361	.156	-3.09**
PERCENTAGE OF IDENTICAL RESPONSE ERROR ON MULTIPLE-CHOICE TESTS	DCNC-IFT	.524	.074	
	DCNC	.286	.141	-2.57*
TIME TAKEN TO COMPLETE PROGRAM (IN MINUTES)	DCNC-IFT	105.88	937.63	
	DCNC	100.52	378.76	-0.738

$^{*}p < .05$
$^{**}p < .01$

TABLE 13

A COMPARISON OF THE DCNC-IFT AND RECALL-CONSTRUCT PROGRAMS ON PROGRAM ERRORS MISSED OR CORRECTED ON IMMEDIATE AND DELAYED CRITERION TESTS

IMMEDIATE TEST	DCNC PROGRAM WITH IDENTICAL MULTIPLE-CHOICE FOILS TEST		RECALL CONSTRUCT PROGRAM COMPLETION TEST	
	Mean	% of Total	Mean	% of Total
POSSIBLE REPEATABLE PROGRAM ERRORS	5.454		7.118	
TEST ITEMS MISSED IDENTICAL ERROR	1.545	28.33%	1.118	15.71%
TEST ITEMS MISSED DIFFERENT ERROR	0.591	10.84%	2.118	29.76%
TEST ITEM RIGHT ERROR ELIMINATED	3.318	60.83%	3.882	54.53%
DELAYED TEST				
POSSIBLE REPEATABLE PROGRAM ERRORS	5.454		7.118	
TEST ITEMS MISSED IDENTICAL ERROR	1.136	20.83%	1.118	15.71%
TEST ITEMS MISSED DIFFERENT ERROR	0.864	15.84%	2.647	37.19%
TEST ITEM RIGHT ERROR ELIMINATED	3.454	63.33%	3.353	47.10%

Note: Comparison may not be entirely valid since the only common denominator is the lead statement in the test items. The DCNC-IFT group took the DCNC multiple-choice program while the other group took the usual linear program. The DCNC-IFT group then took the identical foils multiple-choice format on these items while the RCNC group responded to a completion test.

TABLE 14

A COMPARISON OF THE DCNC-IFT AND DCNC PROGRAMS
ON PROGRAM ERRORS MISSED OR CORRECTED
ON IMMEDIATE AND DELAYED CRITERION TESTS

IMMEDIATE TEST	DCNC PROGRAM WITH IDENTICAL FOILS MULTIPLE-CHOICE TEST		DCNC PROGRAM WITH DIFFERENT FOILS MULTIPLE-CHOICE TEST	
	Mean	% of Total	Mean	% of Total
POSSIBLE REPEATABLE PROGRAM ERRORS	5.826		4.64	
TEST ITEMS MISSED IDENTICAL ERROR	1.652	28.36%	---*	---
TEST ITEMS MISSED DIFFERENT ERROR	0.783	13.44%	0.92	19.83%
TEST ITEMS RIGHT ERROR ELIMINATED	3.391	59.20%	3.72	80.17%
DELAYED TEST				
POSSIBLE REPEATABLE PROGRAM ERRORS	5.826		4.64	
TEST ITEMS MISSED IDENTICAL ERROR	1.608	27.60%	---*	---
TEST ITEMS MISSED DIFFERENT ERROR	0.957	16.43%	1.32	28.45%
TEST ITEMS RIGHT ERROR ELIMINATED	3.261	55.97%	3.32	71.55%

* no possibility in these categories for identical error repetition because test items employed different foils from those used on the corresponding program frames

DCNC, and DCNC-IFT. The findings of interest are that the DCNC condition resulted in the largest percentage of program error corrections from the program to the test and the RCNC (linear) condition the smallest percentage. That is, program errors tend to persist more often under the RCNC condition with the completion test as the criterion measure as compared to the other two conditions with a multiple-choice test as the criterion measure. The small number of cases on which this data is based means that this data is suggestive only.

In addition, the DCNC-IFT condition resulted in a smaller percentage of program error corrections than did the DCNC condition. The use of the same foils on both the program and the multiple-choice criterion test (the DCNC-IFT condition) tends to interfere with the process of error correction.

Efficiency Measures

The question of the efficiency of program conditions has been raised on occasion. It was expected that the condition which included the easy alternatives and the indicated response would require less total program time, which is the case, and unless the program conditions are greatly ineffective in regard to learning one would expect the above conditions to be associated with more favorable efficiency indices.

Time taken to complete the program, mean number right, minutes on the program needed to produce one correct response, number of items learned per minute on the program, and rank order of the program conditions on the basis of the latter measure are presented in Tables 15 and 16. It is noted that the EIC

TABLE 15

MEAN PROGRAM TIME, MEAN NUMBER CORRECT AND RESULTING EFFICIENCY INDICES BASED ON IMMEDIATE AND DELAYED COMPLETION TESTS

Type of Program	Program Time (Minutes)	Mean Number of Test Items Correct	Minutes to Produce One Correct Response	Item Mastery Rate Per Minute	Rank Order
Immediate Completion Test					
DCC	95.56	22.80	4.19	.239	4
DCNC	100.52	19.32	5.20	.192	9
DIC	82.76	29.68	4.00	.249	3
DINC	88.80	18.40	4.83	.207	7
ECC	83.64	19.08	4.38	.228	6
ECNC	77.48	18.20	4.25	.235	5
EIC	71.12	18.44	3.85	.259	1
EINC	69.32	17.52	3.96	.253	2
RCNC	105.18	20.35	5.17	.193	8
Delayed Completion Test					
DCC	95.56	18.00	5.30	.188	5
DCNC	100.52	15.44	6.51	.154	9
DIC	82.76	16.24	5.10	.196	4
DINC	88.80	15.08	5.89	.169	7
ECC	83.64	14.88	5.62	.178	6
ECNC	77.48	15.40	5.03	.199	3
EIC	71.12	16.54	4.30	.232	1
EINC	69.32	15.00	4.62	.216	2
RCNC	105.18	17.65	5.96	.168	8

TABLE 16

MEAN PROGRAM TIME, MEAN NUMBER CORRECT, AND RESULTING EFFICIENCY INDICES BASED ON IMMEDIATE AND DELAYED MULTIPLE-CHOICE TESTS

Type of Program	Program Time (Minutes)	Mean Number of Test Items Correct	Minutes to Produce One Correct Response	Item Mastery Rate Per Minute	Rank Order
Immediate Multiple-Choice Test					
DCC	95.56	26.72	3.57	.280	7
DCNC	100.52	25.08	4.01	.250	8
DIC	82.76	25.64	3.23	.310	4
DINC	88.80	25.24	3.56	.284	6
ECC	83.64	24.96	3.35	.298	5
ECNC	77.48	25.16	3.08	.325	3
EIC	71.12	25.76	2.76	.362	1
EINC	69.32	24.72	2.80	.357	2
RCNC	105.18	25.12	4.18	.239	9
Delayed Multiple-Choice Test					
DCC	95.56	25.72	3.72	.269	7
DCNC	100.52	23.96	4.20	.238	8
DIC	82.76	24.04	3.44	.290	4
DINC	88.80	24.52	3.62	.276	6
ECC	83.64	24.00	3.49	.287	5
ECNC	77.48	23.48	3.20	.303	3
EIC	71.12	23.64	3.01	.332	1
EINC	69.32	22.88	3.03	.330	2
RCNC	105.18	24.88	4.23	.237	9

and EINC are most efficient in every case with the DC and RC group generally showing as least efficient. The major reason for these results is simply that the latter conditions require considerably more time in going through the program, but cannot perform that much better on the criterion tests. The efficiency index, or at least some record of time taken to complete the program, is useful information if two groups should happen to score equally well on the criterion tests. Then the most efficient program would be desirable.

Discussion

Program Errors

As expected, appending the difficult and easy alternatives to the linear program frames resulted in great discrepancies in program error rates with fewer errors being associated with the easy alternatives. More interesting was the fact that (a) constructed responses, as opposed to indicated responses, and (b) the error correction procedure, in comparison to the usual non-correction procedure, both yielded fewer program errors. As was previously mentioned, comparisons between indicated and constructed responses are lacking except in those cases in which the indicated response is confounded with the multiple-choice frame and the constructed response is confounded with the completion frame. But in this comparison in which both groups had to select a response from among a number of alternative responses and then either write or indicate the response selected, writing the response resulted in fewer program errors. This result was not predicted, but one possible explanation of it is that writing the response allowed more time for covert rehearsal of the response. Thus the response was better learned and not so easily missed in the later portions of the program. Perhaps Underwood's (1959) discussion of the relationship between frequency and availability of a response may be appropriately cited as an explanation of the effects of rehearsal.

Finding that the correction procedure resulted in fewer program errors was, in part, expected. Some support for the effectiveness of a procedure which requires making a correct response last in a response sequence would be found in Guthrie's recency principle. As frames are encountered similar to the one in which S missed the response, and corrected, one would expect fewer errors to be made by the Ss making the correction on the earlier frames.

The last significant effect to be reported using program errors as the criterion measure is a two-way interaction between difficulty and correction. This result is explained by noting that the difficult alternatives resulted in a greater number of program errors thus providing an optimum condition for the correction procedure to be effective. It appears that such overt correction procedures will allow for additional rehearsal of the correct response.

Tests as Criterion Measures

None of the studies previously cited found the multiple-choice response to be superior to the constructed response using tests as the criterion measures. (Burton and Goldbeck, 1962; Coulson and Silberman, 1960; Fry, 1960; Hough, 1962; Moore and Smith, 1964; Williams, 1965). The findings of this study, in which the response mode was not separate from the multiple-choice response, did not show the constructed response to be the significant factor when the criterion measure was the immediate completion test constructed of selected frames and part-frames of the program. It was predicted that those Ss required to construct a response would make fewer errors. But the factors of difficulty and correction were both statistically significant, the former at the .01 level of confidence which was not predicted. It should also be recalled that the

RCNC (linear) condition resulted in a larger percent of errors than either of the conditions of difficult alternatives with correction procedure on this criterion measure. Use of the difficult alternatives with the correction procedure may have resulted in greater scrutiny of the stimulus statements of each frame and thus in better performance on the criterion test. It is also possible that the correct response was given somewhat more rehearsal time under these conditions and therefore was better learned.

However, the completion test administered after an approximate two-week delay failed to yield significant differences for the same two factors. The RCNC (linear) condition was found to be almost as effective as the most effective condition (DCC) under analysis. Apparently, the advantage accorded the dimensions of difficulty and correction is temporary. In addition, one can only speculate on the effectiveness of the usual linear program with the added condition of the correction procedure.

The results were exactly the reverse when the multiple-choice test was administered both immediately and approximately two weeks later. None of the factors under study in this investigation showed significant results when the test was administered immediately, but under delayed administration of the same test the dimension of difficulty was significant. However, the factor of difficulty approached significance at the .05 level on the immediate administration of the multiple-choice test. Again it is noted (Table 1) that the RCNC condition showed less loss over time than any of the experimental conditions of this study.

Summary of Data Using Program Errors and Tests as Criterion Measures

Those conditions (the easy alternatives) which were associated with the fewest program errors did not result in superior test performance as compared to those conditions associated with a greater program error rate. However, among the conditions associated with the difficult alternatives the condition which resulted in the least number of program errors also resulted in superior test performance. But the condition which resulted in the greatest number of program errors was the usual linear program which has been designated RCNC. In the light of this evidence the arguments of Skinner (1958) and Porter (1957), that multiple-choice foils would result in error production which in turn would interfere with the learning process, do not seem justified. Instead it would appear that the difficult program alternatives tend to promote inspection behavior of the stimulus section of the frame and perhaps some rehearsal behavior of the response term as well.

Only the difficulty dimension was effective in reducing the number of errors made on the criterion tests of the nature used in this study. The correction procedure was effective in some of those instances when it was used with difficult alternatives. Although the response mode was not a significant factor it is noted that the most effective condition throughout was the DCC condition with the DIC and RCNC conditions next in effectiveness.

Error Analysis

The data on item error repetition and identical response error repetition over the two-week period on both test types are not particularly convincing. Only one factor was statistically significant, that of difficulty when the criterion measure of identical response error repetitions was employed. Further scrutiny of the data showed that the tendency to repeat response errors was associated with the easy alternatives and not the difficult ones as might be expected. This tendency for the test error repetitions to be associated with the easy alternatives may be noted by inspecting the means of the tables in which results were presented.

A convincing explanation for this phenomena is difficult to achieve. It seemed reasonable that the difficult alternatives would tend to result in more program errors (which happened) and thus the tendency to repeat or fixate on these errors should be noted even when test-to-test errors were measured. One problem which complicates the interpretation is that on the completion test there are no alternatives while on the multiple-choice test the alternatvies were in every case different from those employed on the program. How will the program errors affect test-to-test errors under these conditions?

The most logical explanation is that Ss exposed to the difficult alternatives, instead of fixating their responses, actually have more legitimate alternatives to choose from and on this basis, partly by guessing, vary their response from test to test, and perhaps from program to test as well, though this was not ascertained. Those Ss exposed to easy alternatives had been exposed to fewer legitimate alternatives and were not as able to vary their response. The same general statement may be made for the RCNC condition for which a response error repetition percentage was more like that of the Ss exposed to easy alternatives (38.4%). Exposure to difficult and legitimate alternatives may not result in more fixed behavior, but in fact results in more flexible behavior. This result was not found at the expense of a poorer, overall test performance of those Ss exposed to difficult alternatives. It appears that something else may be learned when plausible alternatives were added to the frames which may result in superior transfer behavior as the knowledge is used or applied because of the availability of more information about likely and unlikely responses to a stimulus statement.

Those subjects exposed to easy foils were not required to make discriminations among plausible responses on the program, but were exposed to plausible response alternatives on the criterion test. If they did not know the correct response, then they tended to make the wrong response on some basis of logic that seemed appropriate to them and which process tended to be repeated each time they encountered that particular item. The subjects exposed to the difficult multiple-choice alternatives presumably learned to discriminate among plausible alternatives. If these subjects responded incorrectly they tended to more often vary their response when they next encountered the item because they had been exposed to plausible alternatives before and associated a number of the plausible alternatives with that stimulus statement. It would seem that on the basis of this discussion one could state that those Ss exposed to difficult multiple-choice alternatives had learned more than those exposed to easy alternatives. Whether this is "good" or "bad" is still debatable.

Other Analyses

The comparison between the DCNC and RCNC program was made to evaluate the effect of only the difficult alternatives on Ss performance. The results indicated little difference in the performance of either group exposed to these two program formats except on the criterion of program errors, the greater number of errors being associated with the RCNC program. Appending the difficult alternatives did not add to the effectiveness of the program when tests were the criterion measures. These results tend to agree with a number of other studies previously cited in which no differences were found between multiple-choice formats and the usual linear programs. Little transfer was noted from the program response format to the test response format with Ss exposed to the DCNC and RCNC programs performing equally well on either type of test.

The DCNC and DCNC-IFT comparisons showed significantly more errors and error repetitions being associated with the DCNC-IFT condition on the criterion tests. It is recalled that some of the comparisons were not meaningful as the DCNC group responded to a completion test and the DCNC-IFT group responded to the same test except that the same alternatives used on the program were appended to the completion statements.

It appears that re-presenting the "identical" cue situation on the test as it was encountered in the program results in poorer performance. Encountering the same difficult alternatives on both the program and the test appears to extend the effects of the difficulties in making a choice on the program. That is, the foils encountered on both the program and the test are more familiar because of the prior exposure and it therefore becomes a more difficult task in selecting one of these familiar foils as being correct. Encountering a different set of foils, except for the one which is correct, on the criterion test would then pose a somewhat easier task. The S may more easily recognize the correct response as being correct when it is placed with unfamiliar foils.

A possible explanation of these results is the von Restorff phenomenon (Wallace, 1965) on the superior recall of materials which are in some manner isolated from other materials in which the first set of materials is embedded. For example, if one pair of a list of paired associated is of heterogeneous material and the remaining pairs are of homogeneous material, the heterogeneous pair will be better retained than the remaining pairs. In a sense, prior exposure to one term among four alternatives may tend to isolate that term from the others thus aiding in the recall of the correct response. Then encountering a different set of alternatives, except for the correct response, would aid in selecting the correct response.

It would be of interest to know if Ss responding to the same identical foils on both the program and the test tend to repeat the item errors made on the program to a greater degree than the DCNC group which took a test composed of different difficult foils. Possibly the reappearance of the response they originally chose and missed leads them to select that same response once again, whereas the wrong response is not available fo. the DCNC group so that their choice shifts to the most familiar response which happens to be the correct one. A different research attack would be needed to answer the question of why these results obtained.

The analysis of program errors suggests that repetition of program errors on the test for the DCNC-IFT condition may have accounted for the results just discussed. That is, the DCNC group had a larger percentage of error corrections from program to test than did the DCNC-IFT group. An incidental finding which was earlier reported was that the RCNC condition resulted in the smallest percentage of error correction of items missed on the program and the largest percentage of test items missed with an error different than that which was made on the program. Program errors tend to endure under the DCNC-IFT and RCNC conditions. These results are only tendencies as the data on which the analysis is based are quite meager.

Finally, the matter of efficiency is of concern only if different conditions result in equally good performances, but some conditions require less time in learning the materials. The Difficult-Construct-Correct condition was superior to all other conditions on performance and required less time than two of the program conditions including the usual linear program.

Conclusions and Implications

Conclusions

The major conclusions to be drawn from this study are:

I. Easy alternatives, constructed responses, and error correction resulted in a reduced error rate on the program, but the error correction procedure was effective in reducing errors only when difficult alternatives were employed. The error rate for the usual linear program was higher than when difficult alternatives were appended. This result is in contrast to the belief that using plausible multiple-choice alternatives with a linear program will tend to increase the error rate.

II. Difficult foils and the error correction procedure resulted in superior performance on the completion test administered <u>immediately</u>. None of the factors in this study resulted in differential performance on the <u>delayed</u> completion test. The usual linear program was more effective than any of the conditions except those in which difficult foils and the correction procedure were employed.

III. Only the difficult alternatives affected performance on the multiple-choice test and then only on the delayed test. Only the condition in which difficult foils, constructed response, and the error correction procedure were used was superior to the usual linear program on this criterion measure.

IV. Easy alternatives and the usual linear program promoted identical response error repetitions from the immediate to the delayed test. This result was explained by suggesting that the Ss under these conditions did not have as many plausible alternatives to choose from as did those Ss who were exposed to the difficult alternatives thus tended to repeat their response on both tests.

V. The usual linear program was as effective on a variety of criteria as a modified program to which difficult foils were appended and Ss were required to construct the responses.

VI. A difficult-construct-noncorrect program given to two groups, one which later responded to a test with alternatives identical to those on the program and the other to tests which had difficult foils which were different from those on the program resulted in superior test performance for the latter group. These findings were discussed with perhaps the best explanation being that familiarity with the same set of plausible responses from program to test resulted in more confusion in selecting the correct response than when the correct response is presented with a new set of difficult alternatives.

VII. The efficiency index, a ratio of time spent on the program to correct responses made on the tests is useful only if test performance is approximately equal, and less time is required for one program as compared to the others. Under these conditions the difficult-correct modifications, regardless of mode of response, were the most efficient programs, being somewhat more effective than the usual linear program.

Implications

The results of this study suggest that the use of difficult, plausible alternatives coupled with a correction procedure may be the most effective adaptation of the linear program. Those Ss who took this program out-performed those exposed to the usual linear program on every criterion measure employed. Additional research using the error correction procedure with the typical linear program is needed and would be of interest.

Further study of the effect of these various conditions, plus that of the usual linear program, on the repetition of item and response errors, is essential. There was insufficient data available to study the direct effect of program errors on test errors. The investigator diverted his attention from these data to test-to-test comparisons which are of less interest. Those data which were available on the effect of program errors on test performance were provocative and suggest that plausible alternatives serve a useful function in providing more effective discriminations between the correct response and the incorrect alternatives.

The criterion tests employed in this study were composed of frames and part-frames of the program. This type of test was employed to assess the direct effects of the various conditions of this study. A study of considerable interest could be designed by developing criterion tests to test the subject's ability to transfer his learning under the several program conditions employed in this study with the possible exclusion of the easy alternatives. It was earlier suggested that those subjects exposed to the difficult alternatives learned something different from those subjects exposed to easy alternatives on the usual linear program. Those subjects tended to show more flexible behavior; at least they were less inclined to repeat errors from the immediate to the delayed administration. The question raised here is, "Is this kind of behavior beneficial or detrimental to performance on test situations requiring transfer of learning?" It appears that some interesting questions in regard to program format remained unanswered by the research results presently available.

APPENDICES

Appendix A

An example of the learning program with the difficult multiple-choice alternatives appended. The items here displayed are the first ten items of Set One of the Holland-Skinner program, entitled The Analysis of Behavior.

1 A doctor taps your knee (patellar tendon) with a rubber hammer to test your (A. RECOIL, B. RESPONSES, C. REACTIONS, D. REFLEXES).

Correct here <----

D. REFLEXES

2 If your reflexes are normal, your leg (A. JERKS, B. RESPONDS, C. REACTS, D. RECOILS) with a slight kick to the tap on the leg (the so-called knee-jerk).

Correct here <----

B. RESPONDS

3 In the knee-jerk or patellar-tendon reflex, the kick of the leg is the (A. MOVEMENT, B. RECOIL, C. RESPONSE, D. REFLEX) to the tap on the knee.

Correct here <----

C. RESPONSE

4 The tap delivered by the so-called stimulus object or hammer is the (A. RESPONSE, B. STIMULUS, C. REACTION, D. REFLEX) which elicits the knee-jerk.

Correct here <----

B. STIMULUS

5 The hammer used by the doctor to elicit a knee-jerk is called a (A. OBJECT STIMULUS, B. RESPONSE OBJECT, C. STIMULUS OBJECT, D. STIMULUS RESPONSE).

Correct here <----

C. STIMULUS OBJECT

6 In the knee-jerk reflex, we called the rubber hammer the (A. REFLEX OBJECT, B. ELICITER OBJECT, C. STIMULUS OBJECT, D. RESPONSE OBJECT) and the tap or blow, the (A. CAUSE, B. STIMULUS, C. REFLEX, D. RESPONSE).	
________ ________ Correct here <--- C. STIMULUS OBJECT, B. STIMULUS	
7 An event is explained when its cause is identified. The "cause" or explanation of the knee-jerk is, technically, the (A. STIMULUS, B. RESPONSE, C. REFLEX, D. OBJECT) which elicits it.	
________ ________ Correct here <--- A. STIMULUS	
8 Technically speaking a reflex involves an eliciting stimulus in a process called elicitation. A stimulus (A. STIMULATES, B. CAUSES, C. TRIGGERS, D. ELICITS) a response.	
________ ________ Correct here <--- D. ELICITS	
9 To avoid unwanted nuances of meaning in popular words, we do not say that a stimulus triggers, stimulates, or causes a response, but that it (A. CAUSES, B. ELICITS, C. STIMULATES, D. TRIGGERS) a response.	
________ ________ Correct here <--- B. ELICITS	
10 In a reflex, the stimulus and the elicited response occur in a given temporal order, first the (A. STIMULUS, B. EFFECT, C. CAUSE, D. RESPONSE) and then the (A. CAUSE, B. STIMULUS, C. EFFECT, D. RESPONSE).	
________ ________ Correct here <--- A. STIMULUS, D. RESPONSE	

Appendix B

An example of the learning program with the _easy_ multiple-choice alternatives appended. The items here displayed are the first ten items of Set One of the Holland-Skinner program, entitled _The Analysis of Behavior_.

SET ONE

1	A doctor taps your knee (patellar tendon) with a rubber hammer to test your (A. SOURCES, B. CUES, C. CAUSES, D. REFLEXES).	
D. REFLEXES	Correct here <---	
2	If your reflexes are normal, your leg (A. CONCLUDES, B. RESPONDS, C. ENDS, D. FINISHES) with a slight kick to the tap on the leg (the so-called knee-jerk).	
B. RESPONDS	Correct here <---	
3	In the knee-jerk or patellar-tendon reflex, the kick of the leg is the (A. BEGINNING, B. CUE, C. RESPONSE, D. CAUSE) to the tap on the knee.	
C. RESPONSE	Correct here <---	
4	The tap delivered by the so-called stimulus object or hammer is the (A. OUTCOME, B. STIMULUS, C. RESULT, D. EFFECT) which elicits the knee-jerk.	
B. STIMULUS	Correct here <---	
5	The hammer used by the doctor to elicit a knee-jerk is called a (A. RESULT OBJECT, B. OUTCOME OBJECT, C. STIMULUS OBJECT, D. EFFECT OBJECT).	
C. STIMULUS OBJECT	Correct here <---	

6 In the knee-jerk reflex, we called the rubber hammer the (A. PRODUCT OBJECT, B. RESULT OBJECT, C. STIMULUS OBJECT, D. CONCLUSION OBJECT) and the tap or blow, the (A. RESULT, B. STIMULUS, C. EFFECT, D. OUTCOME).

Correct here <----

C. STIMULUS OBJECT, B. STIMULUS

7 An event is explained when its cause is identified. The "cause" or explanation of the knee-jerk is, technically, the (A. STIMULUS, B. PRODUCT, C. RESULT, D. CONCLUSION) which elicits it.

Correct here <----

A. STIMULUS

8 Technically speaking a reflex involves an eliciting stimulus in a process called elicitation. A stimulus (A. CONCLUDES, B. ENDS, C. FINISHES, D. ELICITS) a response.

Correct here <----

D. ELICITS

9 To avoid unwanted nuances of meaning in popular words, we do not say that a stimulus triggers, stimulates, or causes a response, but that it (A. CONCLUDES, B. ELICITS, C. ENDS, D. FINISHES) a response.

Correct here <----

B. ELICITS

10 In a reflex, the stimulus and the elicited response occur in a given temporal order, first the (A. STIMULUS, B. FINISH, C. OUTCOME, D. RESULT) and then the (A. START, B. CAUSE, C. CUE, D. RESPONSE).

Correct here <----

A. STIMULUS, D. RESPONSE

Appendix C

The criterion test, including both multiple-choice and completion sections.

INSTRUCTIONS: Circle the LETTER of the correct answer.

1. When testing for the presence of a conditioned reflex, it is essential that the ____________ ____________ not be presented on the test trials.

   A. UNCONDITIONED RESPONSE
   B. NEUTRAL STIMULUS
   C. NEUTRAL RESPONSE
   D. UNCONDITIONED STIMULUS

2. An electrically operated food magazine which presents food can be used to reinforce a(n) ____________ of a hungry organism.

   A. ELICITATION
   B. THRESHOLD
   C. LATENCY
   D. RESPONSE

3. When a peck is followed by food, the event is described by saying, "The peck was followed by ______________."

   A. REWARD
   B. REINFORCEMENT
   C. EXTINCTION
   D. CONDITIONING

4. The mother can reinforce an infant's vocal behavior only after at least one vocalization has been ____________.

   A. CONDITIONED
   B. EMITTED
   C. STIMULATED
   D. EXTINGUISHED

5. A process by which a stimulus loses the power to elicit a response is called ______________.

   A. LATENCY
   B. EXTINCTION
   C. FORGETTING
   D. PAIRING

6. The more intense the stimulus, the shorter the ____________ of the reflex.

   A. THRESHOLD
   B. RESPONSE
   C. LATENCY
   D. LENGTH

7. In the knee-jerk reflex, the kick of the leg is the ________ to the tap on the knee.

   A. MOVEMENT
   B. ACTIVITY
   C. REACTION
   D. RESPONSE

8. A response and its eliciting stimulus comprise a(n) ________.

   A. REFLEX
   B. ELICITATION
   C. REACTION
   D. MOVEMENT

9. The magnitude of a response is a function of the ________ of the stimulus which elicits it.

   A. RESPONSE
   B. VELOCITY
   C. INTENSITY
   D. THRESHOLD

10. In ____________, the word "candy" is presented repeatedly without the unconditioned stimulus.

    A. ELICITATION
    B. CONDITIONING
    C. EXTINCTION
    D. REFLEXES

11. An important aspect of the conditioning procedure is the __________, between the presentation of the neutral stimulus and of the unconditioned stimulus.

    A. REFLEX
    B. TIME
    C. THRESHOLD
    D. RESPONSE

12. Reinforcement and behavior occur in the temporal order 1. __________ 2. __________.

    1. A. EXTINCTION
       B. BEHAVIOR
       C. REINFORCEMENT
       D. CONDITIONING

    2. A. BEHAVIOR
       B. EXTINCTION
       C. CONDITIONING
       D. REINFORCEMENT

13. A response elicited by an unconditioned stimulus is a(n) ________ ________.

A. CONDITIONED RESPONSE
B. UNCONDITIONED RESPONSE
C. NEUTRAL STIMULUS
D. UNCONDITIONED REFLEX

14. When a response is elicited by a stimulus without previous conditioning, the sequence is called a(n) ________ ________.

A. UNCONDITIONED REFLEX
B. CONDITIONED RESPONSE
C. CONDITIONED REFLEX
D. CONDITIONED STIMULUS

15. To say that the pigeon will emit pecks at a low rate or frequency is to say that there is a low ________ of pecking the key.

A. CONDITION
B. RESPONSE
C. PROBABILITY
D. THRESHOLD

16. A response elicited by a conditioned stimulus is a(n) ________ ________.

A. NEUTRAL STIMULUS
B. UNCONDITIONED STIMULUS
C. CONDITIONED RESPONSE
D. CONDITIONED REFLEX

17. If a child is to learn to salivate to the word "candy," "candy" and eating candy must occur nearly ________.

A. CONCURRENTLY
B. TOGETHER
C. JOINTLY
D. SIMULTANEOUSLY

18. In the unconditioned hand-withdrawal reflex, the movement of the arm is the ________ ________.

A. UNCONDITIONED RESPONSE
B. UNCONDITIONED STIMULUS
C. UNCONDITIONED REFLEX
D. CONDITIONED REFLEX

19. The more intense the stimulus, the greater the __________ of the response

A. THRESHOLD
B. MAGNITUDE
C. LATENCY
D. POWER

20. Candy put in the mouth of the child for the first time __________ salivation.

A. EVOKES
B. CAUSES
C. TRIGGERS
D. ELICITS

21. The response is not emitted in the process called __________.

A. ELICITATION
B. LATENCY
C. CONDITIONING
D. FORGETTING

22. Reinforcing a response produces an increase in the __________ that a response will occur again.

A. REINFORCEMENT
B. THRESHOLD
C. PROBABILITY
D. CONDITIONING

23. In the conditioning experiment, the __________ of the conditioned response fluctuated slightly between 1 and 2 seconds, remaining essentially constant after Trial 30.

A. LATENCY
B. EXTINCTION
C. THRESHOLD
D. INTENSITY

24. A kick of the leg is __________ by a tap on the patellar tendon.

A. ELICITED
B. EMITTED
C. INSTIGATED
D. CAUSED

25. An initially neutral stimulus ceases to be able to elicit a response after __________ has taken place.

A. LATENCY
B. UNCONDITIONING
C. ELICITATION
D. EXTINCTION

26. As the number of trials in which a conditioned stimulus is presented alone increases, the (1) __________ of a conditioned response decreases until (2) __________ is complete.

(1) A. EXTINCTION
B. MAGNITUDE
C. CONDITIONING
D. THRESHOLD

(2) A. LATENCY
B. LEARNING
C. UNCONDITIONING
D. EXTINCTION

27. A given peck on the key is a(n) __________.

A. STIMULUS
B. THRESHOLD
C. PROBABILITY
D. RESPONSE

INSTRUCTIONS: Fill in the blank with the correct word or words to complete the statement.

1. In the unconditioned hand-withdrawal reflex, heat is the ________ ________.

2. When a peck is followed by food, the food is called a(n) ________ ________.

3. The room used for conditioning experiments is designed to ________ uncontrolled factors which might affect the experimental result.

4. In reflex behavior, the process by which a new stimulus comes to elicit a response is called ________.

5. As the number of trials in which a conditioned stimulus is presented alone increases, the ________ of a conditioned reflex increases.

6. In a reflex, a sufficient explanation of the response is a description of the preceding ________.

7. Food is not reinforcing, unless the animal has first been ________ ________ food for some time.

8. As the number of pairings of the conditioned and unconditioned stimuli increases, the ________ of the conditioned reflex decreases until it reaches a limit.

9. Since a tone has no effect on salivation before conditioning, it is a(n) ________ ________.

10. As the number of pairings of the conditioned and unconditioned stimuli increases, the ________ of the conditioned response increases until it reaches a limit.

11. The softest touch on the surface of the eye needed to elicit a blink marks the ________ of the stimulus.

12. A stimulus able to elicit a response only after conditioning is called a(n) ________ ________.

13. A tone and the salivation elicited comprise a(n) ________ (salivary) ________.

14. If the conditioning experiment described in Set 4 had continued to Trial 50, the ________ of the conditioned response would probably have been in the vicinity of 60.

15. A stimulus able to elicit a response without previous conditioning is called a(n) __________ __________.

16. The general behavior of "pecking the key" is a(n) __________.

17. __________ behavior is influenced by the consequences of previous, similar responses.

18. A new neutral stimulus is able to elicit a response after __________ has taken place.

19. Since no stimuli are observed for such responses as flicking leaves or bar pressing, we cannot say that these responses are __________ by stimuli, as are the responses in reflexes.

20. In the knee-jerk reflex, we call the tap or blow the __________.

21. To condition a reflex, an initially __________ __________ is paired with a(n) __________ __________.

22. In __________ behavior, a stimulus precedes the response.

23. In a reflex, the stimulus must be intense enough to exceed the __________ or no response will occur.

24. The response of turning on the electrically-operated heat lamp will be __________ more frequently in the future if the organism is cold and hungry.

25. The time between the tap and the kick is the __________ of the knee-jerk reflex.

26. In the knee-jerk reflex, we call the rubber hammer the __________ __________.

27. The response is emitted without reinforcement only in the process called __________.

BIBLIOGRAPHY

Alter, Millicent & Silverman, R. E. The response in programed instruction. J. programed Instruct., 1963, 1, 55-79.

Burke, C. J., Estes, W. K., & Hellyer, S. Rate of verbal conditioning in relation to stimulus variability. J. exp. Psychol., 1954, 48, 153-161.

Burton, B. B. & Goldbeck, R. A. The effect of response characteristics and multiple-choice alternatives on learning during programed instruction. Technical Report No. 4, San Mateo, Calif.: American Institute for Research, 1962. 16 pp.

Coulson, J. E. & Silberman, H. G. Effects of three variables in a teaching machine. J. educ. Psychol., 1960, 51, 135-143.

Fry, E. B. A study of teaching-machine response modes. In A. A. Lumsdaine & R. Glaser (Eds.), Teaching machines and programmed learning: A source book. Washington, D. C.: National Education Association, 1960, Pp. 469-474.

Glaser, R. Toward a behavioral science base for instructional design. In R. Glaser (Ed.) Teaching machines and programed learning, II: Data and directions. Washington, D. C.: National Education Association, 1965. Pp. 771-809.

Holland, J. G. Research on programing variables. In R. Glaser (Ed.) Teaching machines and programed learning, II: Data and directions. Washington, D. C.: National Education Association, 1965. Pp. 66-117.

Holland, J. G. & Skinner, B. F. The Analysis of Behavior, New York: McGraw-Hill Book Company, Inc., 1961.

Hough, J. B. An analysis of the efficiency and effectiveness of selected aspects of machine instruction. J. educ. Res., 1962, 55, 467-471.

Irion, A. L. & Briggs, L. H. Learning task and mode of operation variables in use of subject matter trainer. AF PTRC-TR-57-8, October 1957.

Kaess, W. & Zeaman, D. Positive and negative knowledge of results on a Pressey-type punchboard. J. exp. Psychol., 1960, 60, 12-17.

Lindquist, E. F. Design and analysis of experiments in psychology and education, Boston: Houghton Mifflin Co., 1953.

Moore, W. J. & Smith, W. I. Role of knowledge of results in programed instruction. Psychol. Rep., 1964, 407-423.

Porter, D. A. Critical review of a portion of the literature on teaching devices. Harv. educ. Rev., 1957, 27, 126-147.

Pressey, S. L. Basic unresolved teaching-machine problems. Theory into Practice, 1962, 1, 30-37.

Scheffe, H. The analysis of variance, New York: Wiley, 1959.

Skinner, B. F. Teaching Machines. Science, 1958, 128, 969-977.

Suppes, P. & Ginsberg, Rose. Application of a stimulus sampling model to children's concept formation with and without overt correction response. J. exp. Psychol., 1962, 63, 330-336.

Underwood, B. J. Verbal learning in the educative process. Harv. educ. Rev., 1959, 39, 107-117.

Wallace, W. P. Review of the historical, empirical, and theoretical status of the von Restorff phenomenon. Psychol. Bull., 1965, 63, 410-424.

Williams, Joanna P. Effectiveness of constructed response and multiple choice programing modes as a function of test mode. J. educ. Psychol., 1965, 56, 111-117.